Information-theoretic properties of auditory sequences dynamically influence expectation and memory

Kat Agres[1]

Samer Abdallah[2]

Marcus Pearce[1]

1. Queen Mary University of London, School of Electronic Engineering and Computer Science
2. University College London, Department of Computer Science

Keywords: expectation; recognition memory; predictive coding; information theory; computational modeling; auditory perception; music cognition

Author Note

A preliminary version of this research was previously reported in Agres, Abdallah, and Pearce (2013), in the conference proceedings of the Cognitive Science Society. The research was supported by EPSRC via grant number EP/D038855/1: *Neural and Information Dynamics in the Processing of Musical Structure.*

Address for correspondence:
Kat Agres
School of Electronic Engineering & Computer Science
Queen Mary, University of London
London E1 4NS, UK
Tel: +44 (0)77 5835 6241
Email: kathleen.agres@qmul.ac.uk

Marcus Pearce
School of Electronic Engineering & Computer Science
Queen Mary, University of London
London E1 4NS, UK
Tel: +44 (0)20 7882 6207
Email: marcus.pearce@qmul.ac.uk

Abstract

A basic function of cognition is to detect regularities in sensory input to facilitate the prediction and recognition of future events. It has been proposed that these implicit expectations arise from an internal predictive coding model, based on knowledge acquired through processes such as statistical learning, but it is unclear how different types of statistical information affect listeners' *memory* for auditory stimuli. We used a combination of behavioral and computational methods to investigate memory for non-linguistic auditory sequences. Participants repeatedly heard tone sequences varying systematically in their information-theoretic properties. Expectedness ratings of tones were collected during three listening sessions, and a recognition memory test was given after each session. Information-theoretic measures of sequential predictability significantly influenced listeners' expectedness ratings, and variations in these properties had a significant impact on memory performance. Predictable sequences yielded increasingly better memory performance with increasing exposure. Computational simulations using a probabilistic model of auditory expectation suggest that listeners dynamically formed a new, and increasingly accurate, implicit cognitive model of the information-theoretic structure of the sequences throughout the experimental session.

1. Introduction

Effective perceptual systems must learn and remember regularities in sensory input, so as to generate accurate expectations for future events. Expectation and prediction are thought to be important mechanisms in many areas of cognition including language processing (Cristia, McGuire, Seidl & Francis, 2011; DeLong, Urbach & Kutas, 2005; Hale, 2006; Levy, 2008; Saffran, 2003a, Saffran, 2003b), visual perception (Bar, 2007; Bubic, von Cramon & Schubotz, 2010; Egner, Monti & Summerfield, 2010), music perception (Huron, 2006; Meyer, 1956; Pearce, 2005; Pearce & Wiggins, 2012), and motor sequencing (Wolpert & Flanagan, 2001). Recent accounts of the role of prediction in cognitive and neural processing of sensory information suggest that expectations about future events come from the experience and prediction of past events. In particular, from the perspective of hierarchical *predictive coding* (Clark, 2013; Friston, 2010; Friston & Kiebel, 2009), an internal model of the sensory environment compares top-down predictions about the future with the actual events that transpire, and error signals generated from the comparison drive learning to improve future predictions. These prediction errors occur at a series of hierarchical levels, each reflecting integration of information over successively longer time-scales. The idea that top-down predictions play a central role in constructing coherent representations of incoming sensory input has a venerable history (Barlow, 1959; Dayan, Hinton, Neal & Zemel; 1995; Gregory, 1980; Helmholtz, 1866). However, recent years have seen a resurgence of interest in cognitive mechanisms of statistical and probabilistic learning that are thought to underlie the generation of expectations in these accounts. In the auditory modality, these expectations relate especially to information contained in structured pitch sequences, and evidence supports hierarchical predictive coding of pitch perception for such sequences (Furl, Kumar, Alter, Durrant, Shawe-Taylor & Griffiths, 2011; Kumar, et al., 2011).

1.1. Statistical Learning and Predictive Processing of Sequential Structure

Perceptual surprise resulting from a mismatch between sensory information and top-down predictions can be minimized by updating those predictions through dynamic statistical learning to more accurately reflect the sequential structure of the sensory signal. Statistical learning has been studied primarily by assessing the ability to segment isochronous sequences (i.e., sequences with no variation in duration or time interval between the onset of consecutive events) following exposure to stimuli with known statistical structure. Saffran, Newport, and Aslin (1996b) found that after several minutes of exposure to artificially constructed isochronous syllable sequences, adults are able to accurately segment 3-syllable 'words' distinguished only by having lower syllable transition probabilities between compared to within the words. Following exposure, participants are presented with pairs consisting of a valid word (where the three syllables occur between word boundaries) and a non-word (where a word boundary occurs within the three syllables) and asked to identify the one that is most familiar. Performance is typically above chance, demonstrating that participants were able to accurately segment the syllable streams using the underlying statistics defining word boundaries.

Subsequent research found that 8-month-old infants also perform above chance on this task (Saffran, Aslin & Newport, 1996a). Subsequent work using this paradigm has demonstrated sensitivity to statistical properties of tone sequences (Saffran, Johnson, Aslin & Newport, 1999), and pitch interval sequences (Saffran, Reeck, Niebuhr & Wilson, 2005; Saffran & Griepentrog, 2001). Research on the influence of statistical and implicit learning on language acquisition has tended to focus on segmentation and chunking of language at different levels of hierarchical organization (Perruchet & Pacton, 2006; Romberg & Saffran, 2010; Mirman, Magnuson, Graf Estes & Dixon, 2010) and acquisition of syntactic categories (Redington, Chater & Finch, 1998).

These results demonstrate an ability to learn the statistical structure of unfamiliar stimuli. In the present research, we focus on understanding the basic cognitive processes of expectation and memory and how they are influenced by statistical learning. Expectation has been extensively studied in research on music perception, which suggests that listeners implicitly acquire knowledge about the statistical structure of music and that this knowledge guides their perception of subsequent music (Huron, 2006; Krumhansl, 1990; Pearce & Wiggins, 2006; Tilmann, 2012; Rohrmeier & Rebuschat, 2012). Furthermore, there is evidence that expectations are influenced in this way both through long-term exposure to music (Krumhansl, 1990), and through learning the properties of the local context (Oram & Cuddy, 1995; Tillmann, Bharucha, & Bigand, 2000; Tillmann, Bigand, & Pineau, 1998). This learning has been conceptualized as the acquisition of an internal representation of the statistical properties of the musical sequences to which listeners are exposed over a range of temporal scales (Krumhansl & Kessler, 1982; Krumhansl, 1990; Temperley, 2007). This process of statistical learning allows listeners to generate probabilistic predictions about forthcoming musical events, dependent on the prior musical context and previously acquired schematic expectations for the musical style in question (Krumhansl, Louhivuori, Toiviainen, Järvinen, & Eerola, 1999; Pearce, Ruiz, Kapasi, Wiggins, & Bhattacharya, 2010; Pearce & Wiggins, 2006). Research to date has not examined how statistical properties influence expectations and recognition memory over periods of increasing exposure to stylistically unfamiliar auditory tone sequences.

1.2. Information-theoretic Accounts of Auditory Processing

Most research on statistical learning assumes that listeners acquire simple cognitive models of statistical structure, corresponding to first-order Markov transition tables. This leaves open the questions of exactly how these models are acquired and how they are used to estimate the predictability of entire sequences and events within them. Information theory

provides a way of describing and quantifying, in precise terms, the information contained in a signal. This is especially useful for clarifying how cognitive systems process and learn temporal sensory signals; and indeed, information-theoretic measures such as *information content,* a measure of surprisal, and *entropy*, a measure of uncertainty, have been used to simulate successfully anticipation of forthcoming sensory input, such as music and language (e.g., Elman, 1990; Brent, 1999; Manning & Schütze, 1999; Hale, 2006; Levy, 2008; Abdallah & Plumbley, 2009; Pearce, 2005; Hansen & Pearce, 2014).

There is a long history of interest in information theoretic models of non-linguistic auditory sequences (e.g., Ames, 1989; Cohen, 1962; Knoppoff & Hutchinson, 1981, 1983; Moles, 1966; Youngblood, 1958), but many approaches suffered from using simple predictive models, small datasets and inflexible representations (see reviews by Cohen, 1962, and Ames, 1989). Furthermore, they focused on computing and comparing the information-theoretic properties of entire corpora, rather than building dynamic predictive models that learn incrementally through exposure (Cohen, 1962; Pearce & Wiggins, 2012). In recent years, sophisticated, dynamic probabilistic models such as IDyOM (Pearce, 2005; see also Section 3.1) have successfully derived information-theoretic properties of auditory sequences that accurately account for listeners' expectations in many listening tasks (Egermann, Pearce, Wiggins, & McAdams, 2013; Omigie, Pearce & Stewart. 2012; Omigie, Pearce, Williamson, & Stewart, 2013; Pearce, 2005; Pearce et al., 2010; Hansen & Pearce, 2014). Recent research has also developed more sophisticated information-theoretic measures that systematically distinguish different ways in which a stimulus can be unpredictable (Abdallah & Plumbley, 2009; Abdallah & Plumbley, 2010, Abdallah & Plumbley, 2012).

Research to date has not applied these recently-developed models and information-theoretic measures to learning and memory for stylistically unfamiliar auditory stimuli over periods of increasing exposure.

1.3. Memory in the Cognitive Processing of Auditory Sequences

Statistical learning of sequential structure in auditory perception requires that listeners form in memory some record of the frequency with which different auditory events appear in different sequential contexts. Much attention has focused on elucidating the relationship between complexity and memory, and comparing the extraction of rules compared with learning particular exemplars from sets of auditory stimuli. In music, for example, research has shown that recognition memory for melodies can be influenced by the complexity of motifs (brief, recurring passages of music that hold thematic meaning) and melodic distinctiveness (Müllensiefen & Halpern, 2014), familiarity (e.g., presenting a well-known tune) (Bartlett, Halpern, & Dowling, 1995), and listener traits (e.g., age and experience) (Dowling, Bartlett, Halpern, & Andrews, 2008). Predominantly, unfamiliar and more complex stimuli often yield poor explicit recognition memory (Halpern & Bartlett, 2010), suggesting a relationship between predictability and recognition memory.

Research on implicit sequence learning has also explicitly connected the statistical structure of stimuli with recognition of particular stimuli. For example, evidence suggests that the repetition of a small number of stimulus exemplars may lead to satisfactory recognition of those exemplars but unsuccessful internalization of the statistical rules (indicated by generalization to new exemplars), while repetition of a larger number of exemplars can lead to better generalization but worse recognition performance (Loui & Wessel, 2008; Cleeremans, Destrebecqz, & Boyer, 1998). In other words, the statistical structure of auditory sequences may be extracted while the particular exemplars themselves are not retained. While these findings clarify the interplay between repetition and learning, research to date has not examined the information-theoretic properties of the stimulus in great detail, or how these properties influence statistical learning and memory over periods of increasing exposure.

1.4 The present research

In the present research, we investigate how expectation and recognition memory change with repeated exposure to stylistically unfamiliar auditory tone sequences that vary systematically in predictability (defined in precise, information-theoretic terms). We use carefully controlled, artificially-constructed, non-linguistic auditory stimuli so as to focus specifically on the effects of information-theoretic properties of stimulus structure on expectation and memory in domain-general sequence processing (without interference from explicit referential semantics, for example, associated with linguistic stimuli). Probabilistic models are subsequently employed to simulate the cognitive process of online statistical learning. Clark (2013, p. 8) has argued that "the nervous system is fundamentally adapted to deal with uncertainty, noise, and ambiguity, and that it requires some (perhaps several) concrete means of internally representing uncertainty." Using carefully constructed tone sequences, we examine different information-theoretic representations of stimulus uncertainty and unpredictability. Through behavioral testing and computational simulation using these sequences, the present research aims to elucidate the underlying cognitive probabilistic models that listeners develop through statistical learning, and how these models have an impact on the expectedness of individual tones, as well as memory for particular tone sequences, over a period of increasing exposure.

In this paper, for clarity of expression, we use *expectation* and *expectedness* to refer to human cognitive processes and *prediction* and *predictability* to refer to computational simulations. Expectation and prediction refer to the general process of anticipating future events. Expectedness and unexpectedness refer to the subjective likelihood (or surprisal) of a particular perceived event for a listener. Predictability and unpredictability refer to the likelihood of an event or, more often, a sequence of events according to a computational

model (the only exception arising during discussions of predictive coding, which has been applied both to computational models and human cognitive processes).

2. Behavioral Study

This study uses a *generative probabilistic model* to create a set of tone sequences varying systematically across three information-theoretic measures. Varying the sequences' statistical structure allows us to assess which properties have the greatest impact on listeners' auditory expectations and memory for tone sequences. We focus on testing the relative influence of three factors based on the information-theoretic concepts of *Entropy Rate, Multi-Information Rate*[1], and *Predictive Information Rate* (see Abdallah & Plumbley, 2009; Abdallah & Plumbley, 2010, Abdallah & Plumbley, 2012). These measures, discussed in detail below, are defined for random processes with a known probability distribution.

Listeners cannot know the probability distribution used by the generative model *a priori* but they can estimate it by listening to tone sequences generated from the distribution. Therefore, rather than using the generative model to derive information-theoretic properties of the generated stimulus sequences, we use a corresponding *analytical model*, which assumes that notes are sampled from a Markov process with an unknown transition matrix, and estimates the transition matrix from the observed tones within each sequence. The analytical model therefore processes each stimulus sequence in a sequential event-by-event manner, dynamically updating its estimated probability model as it does so.

Both the generative and analytical models are first-order n-gram models supplied with the same initial Bayesian prior consisting of a transition matrix derived from a large corpus of Western tonal melodies in a Major key (see Section 2.2.2). This ensures that the analytical model never encounters a previously unseen tone (thus avoiding the zero-frequency problem,

[1] Referred to as *Redundancy* in Abdallah and Plumbley (2009).

see Section 3.1). Although the prior was intended to lend the stimulus sequences some familiar tonal musical structure, the tone distributions across stimuli did not correlate with tone profiles derived from the experiments of Krumhansl & Kessler (1982; see also Krumhansl, 1990), well-known measures of tonal structure (see Section 2.2.2). Further, the computational simulations of listeners' expectation and memory performance were not improved by adding Western music training to the model, suggesting that listeners did not process the stimuli as familiar, tonal melodies (see the Appendix). Therefore, in the remainder of the paper we treat the stimuli as non-musical.

The analytical model yields the three *pointwise* information measures examined in the present research, *Information Content*, *Coding Gain*, and *Predictive Information*, which are defined for particular events in particular sequences. They correspond to the static information rates used to generate the sequences, which are defined for random processes with known distributions. We now describe the measures in detail (see also Table 1).

- **Information Content,** corresponding to Entropy Rate, is a measure of the unexpectedness of an event in a sequence given the previous event. At any integer time t, let x_t be the note occurring at that time, and Θ_t be the estimated transition matrix using information available at the previous timestep $t\text{-}1$. The model, Θ_t, generates a conditional probability distribution governing the identity of x_t, given x_{t-1}: $p(x_t|x_{t-1}, \Theta_t)$. The Information Content at time t is the negative log probability of x_t given the context and the estimated model: $-\log p(x_t|x_{t-1}, \Theta_t)$. Entropy is the Information Content averaged over all possible observations at a given point in the sequence, while the Entropy Rate of the random process (see above) is the entropy averaged over all possible contexts (Abdallah & Plumbley, 2009).
- **Coding Gain**, corresponding to Multi-Information Rate, measures how different the information content of the current event would be if the model didn't know the

identity of the previous event (in other words, how much information about the current event the model gains from the previous event). Coding Gain at time t quantifies how much the model's ability to predict the current observation depends on having observed the preceding observations, and is a difference of log probabilities: log $p(x_t|x_{t-1}, \Theta_t)$ - log $p(x_t|\Theta_t)$, where the latter term is derived from the stationary distribution of the transition matrix. The Multi-Information Rate is the Coding Gain averaged over all possible observations (x_t) and contexts (x_{t-1}). This is equivalent to the mutual information (Mackay, 2003) between x_t and x_{t-1} computed from their joint distribution, $p(x_t, x_{t-1} | \Theta_t)$ (Abdallah & Plumbley, 2009).

- **Predictive Information**, corresponding to Predictive Information Rate, quantifies how much the current event improves precision in predicting the next event. Predictive Information is quantified as the Kullback-Liebler divergence (Mackay, 2003) between two probability distributions: $DKL(p(x_{t+1}|x_t,\Theta_{t+1})||p(x_{t+1}|x_{t-1},\Theta_t))$, representing the observer's probabilistic beliefs about x_{t+1} before and after the observation of x_t. Predictive Information Rate[2] is the Predictive Information averaged over all possible observations (x_t) and contexts (x_{t-1}), which is equivalent to a conditional mutual information (Mackay 2003) between x_t and x_{t+1} given x_{t-1} according to their joint (trivariate) distribution, $p(x_t, x_{t+1}, x_{t-1})$.

We use Entropy Rate, Multi-Information Rate and Predictive Information Rate from the generative model to select distributions to create the stimulus sequences. We then use Information Content, Coding Gain and Predictive Information from the analytical model to measure the information-theoretic properties of individual events in the stimulus sequences, and then average across every event in a sequence to compute *sequence measures,* representing the overall predictability of entire sequences.

[2] Referred to as *instantaneous predictive information* in Abdallah and Plumbley (2009).

----------Insert Table 1 about here----------

To investigate the processes underlying auditory expectation and memory, listeners were exposed to isochronous tone sequences produced by the generative model over three listening sessions. In each listening session, participants heard sequences and rated the expectedness of a tone (termed the *probe tone*) within each sequence. Probe tones varied in terms of information content (representing unexpectedness) across sequences. We focus on the Information Content of the probe tone because it is a straightforward measure of unexpectedness that accounts well for listeners' expectations (e.g., Hansen & Pearce, 2014; Pearce et al., 2010). A recognition memory test containing old and new sequences followed each listening session. The timing of tones was not experimentally manipulated, as we sought to constrain expectation and memory mechanisms to pitch relationships between tones, controlling for potential confounding effects of temporal structure. This experimental design enabled us to compute information-theoretic measures for every tone sequence, and compare the effect of these information-theoretic properties on probe tone expectedness ratings in the listening sessions as well as recognition memory performance in the test sessions.

2.1 Hypotheses

We propose specific hypotheses about the effects of the information-theoretic properties of the stimuli on the expectedness of individual tones and recognition memory for tone sequences.

2.1.1. Expectedness

Based on the findings reviewed above, we hypothesize that, for individual events in the sequences, high Information Content probe tones will produce greater unexpectedness than low Information Content tones. We also hypothesize that expectedness of tones will increase with greater exposure across the three listening sessions as listeners form an

increasingly accurate cognitive model of the statistical structure of the stimuli. We also envisage that the context of the probe tone (that is, the statistical properties of the sequence in which the tone is embedded) will influence the perceived expectedness of the tone. Specifically we hypothesize that:

- The hypothesized relationship between information content and expectedness for tones will hold when those tones are embedded in predictable sequences, but
- unpredictable sequences will confound the generation of expectations leading to moderate expectations for both high and low information content events.

The underlying assumption here, following the predictive coding framework (Clark, 2013; Friston, 2010), is that successful prediction depends on having a structured context with which to generate coherent expectations. As such, the "weight given to sensory prediction error is varied according to how reliable (how noisy, certain, or uncertain) the signal is taken to be" (Clark, 2013, p. 10). In other words, the precision and strength of expectations should reflect the predictability of the signal itself. Unpredictable contexts are likely to generate uncertain expectations characterized by high entropy in which every possible next event is equally likely (Hansen & Pearce, 2014). In this research, we use the sequence measures defined above (Information Content, Coding Gain and Predictive Information) as operational definitions of sequence predictability. Highly unpredictable sequences are those with high Sequence Information Content, high Sequence Predictive Information, and low Sequence Coding Gain, and vice versa for predictable sequences.

2.1.2. Recognition memory

We hypothesize that unpredictable sequences will be more difficult to encode, and therefore yield worse memory performance and lower confidence ratings than predictable

sequences. Highly unpredictable sequences are those with high Sequence Information Content, high Sequence Predictive Information, and low Sequence Coding Gain, and vice versa for predictable sequences.

We further hypothesize that predictable sequences conforming to listeners' expectations will be easier to encode (Atkinson & Shiffrin, 1968). It is also possible, however, that unpredictable sequences, while difficult to encode in memory, will sound more distinctive than predictable sequences, thereby facilitating accurate discrimination of familiar from unfamiliar stimuli in recognition test performance. Considering the trade off between these two conflicting effects, we hypothesize that the former will prove more influential in recalling previously heard tone sequences because the potential distinctiveness of unpredictable stimuli will be of little use if those stimuli are difficult to encode in the first place. It may be the case, however, that new unpredictable sequences which have not been previously encoded, will prove more distinctive and easier to reject in the recognition memory task (cf. Müllensiefen & Halpern, 2014).

Lastly, because every tone sequence is presented in each of the three listening sessions, we also aim to clarify the learning trajectory of the different classes of tone sequences; that is, how auditory information represented in short-term memory gradually becomes more richly encoded in long-term memory, and how sequential predictability, represented by the information-theoretic measures outlined above, influences this process over time. We hypothesize that more accurate memory performance, and increasing confidence of recognition judgments, will arise from increased exposure to the sequences.

2.1.3. Differences between the information-theoretic measures

A further goal of this research is to examine whether listeners show different levels of sensitivity to the different information-theoretic measures of unpredictability. Although there is no previous empirical research on which to base a hypothesis, the three measures described

above capture different kinds of unpredictability that may vary in the extent to which they capture aspects of listeners' task performance. Information content is a simple measure of the predictability of each tone in a sequence, given the preceding tone. Coding Gain, on the other hand, reflects the degree to which a tone is statistically dependent on its predecessor in the sequence, i.e., the extent to which knowing the previous tone increases the predictability of the next tone. While these measures are derived from the probability of particular events, Predictive Information differs in that it is derived from predictive probability *distributions*. This measure reflects the divergence between the distribution of a model that sees the current event and one that does not (see definition above). It therefore measures the extent to which the current event influences the model observer's uncertainty about future events, i.e., a property of the observer rather than of the sequential events per se.

2.2. Method

2.2.1. Participants

Twenty-three students (12 female and 11 male; mean age = 21.0) at Cornell University participated in this study for extra credit in a psychology course. The participants had an average of 1.61 years (*SD* = 1.88) playing music in the previous five years, and an average of 5.82 years (*SD* = 4.54) of lifetime experience playing an instrument.

2.2.2. Stimuli

The 24 sequences in the listening sessions and 24 New sequences in the test sessions (48 stimuli in total) each comprised 24 isochronous tones, played in a piano timbre. Each tone was 500 ms in duration, yielding sequences that were 12-seconds-long each. The sequences were generated using an alphabet of 7 pitches (representing one octave of the major diatonic scale starting at C_4). To construct the tone sequences, many transition matrices were generated randomly using a product of first-order Dirichlet distributions (Bertuccelli &

How, 2008), one for each column of the transition matrix. From each matrix, one sequence of 24 notes was sampled. A subset of these was then selected manually to ensure a good spread in the 3-dimensional information space formed by the distribution measures described above: Entropy Rate, Multi-Information Rate, and Predictive Information Rate (see Fig. 1).

----------Insert Figure 1 about here----------

The Dirichlet distributions were biased towards a tonal transition matrix derived from the scale degrees of Canadian folk songs/ballads, Chorale melodies, and German folk songs in a major key (the same corpus described in Table 2 of Pearce & Wiggins, 2006). All stimuli were generated using the notes of the diatonic scale of C Major, to ensure that the sequences were consistent in terms of tonality. Nonetheless, across all stimuli, the zeroth-order distribution for the seven pitches does not show significant correlations with the Krumhansl and Kessler (1982) tone profiles for keys in major, $r(5) = -.14$, $p = .77$, or minor, $r(5) = .19$, $p = .68$. For examples of the stimuli, please refer to Fig. 2.[3]

----------Insert Figure 2 about here----------

2.2.3. Procedure

After receiving written and verbal instructions, participants listened to tone sequences in three sessions, each lasting approximately 15 minutes and followed by a brief test session. In the listening sessions, participants heard each of the 24 tone sequences (presented in a different order in each session). On each trial, participants were asked to rate the expectedness of a particular tone (the probe tone) within the sequence. This tone was identified visually on the computer screen via a clock counting down on the subsequent tones of the sequence (Pearce et al., 2010). When the clock returned to midnight, participants rated the expectedness of the concurrently sounding tone on a scale from 1 to 5, where '1'

[3] The complete set of stimuli created for this study is available online at http://webprojects.eecs.qmul.ac.uk/marcusp/software/AgresAbdallahPearceStimuli.zip.

represented highly unexpected and '5' represented highly expected. To help clarify the concept of expectedness, the experimenter sang the pitches of a diatonic major scale and drew attention to the expectation the listener may have experienced for the culmination (octave) of the scale.

Probe tones could occur between tones 17 and 23 of each melody. This was to ensure that listeners had sufficient exposure to each sequence prior to making an expectedness rating. The probe tone never occurred on the last tone to avoid a possible confound of perceptual closure.

Each listening session was followed by a test session. Sixteen test stimuli were presented in each of the three test sessions, where 8 sequences were *Old* (had been presented previously) and 8 were *New*. After each test sequence, participants responded "Yes" or "No" to whether they had heard the sequence before. Upon responding, the listener made a confidence rating on a scale from 1 to 5 where '1' represented "not confident" and '5' represented "very confident".

A distinct 500 ms white noise clip was played after every tone sequence in the listening and test sessions to perceptually "reset" echoic memory and ensure that expectedness ratings and memory judgments were based only on the current trial. The study was administered on a MacBook Pro laptop, and stimuli were presented and responses collected using Psychophysics Toolbox (Version 3) within the programming environment of MATLAB 2010a (MathWorks, Inc). Participants listened to stimuli over headphones set to a comfortable listening volume.

2.3. Results

2.3.1. Expectedness Ratings during Listening Sessions

To examine which factors in the listening sessions had the greatest impact on expectation, a stepwise regression was performed to select the variables to include in a

logistic regression analysis. This approach was taken to address potential multicollinearities between the information-theoretic factors. The chosen factors for the logistic regression were *Probe Tone Information Content*, *Sequence Information Content*, *Sequence Coding Gain*, *Sequence Predictive Information*, and *Listening Session*, with *Average Expectedness Ratings* as the dependent measure. This and the subsequent stepwise regressions used minimum AIC as the stopping rule.

As hypothesized, there was a significant main effect of *Probe Tone Information Content* ($F(1, 64) = 117.13$, $p < .001$), with high Information Content tones rated as less expected (see Fig. 3). In addition to this main effect, there were significant interactions between Probe Tone Information Content and two whole sequence measures: *Probe Tone Information Content* X *Sequence Predictive Information* ($F(1, 64) = 37.47$, $p < .001$), and *Probe Tone Information Content* X *Sequence Coding Gain* ($F(1, 64) = 8.43$, $p < .01$). The sign of the model coefficients for these interactions (1.20 and -0.18 respectively) indicate that the relationship between Probe Tone Information Content and Expectedness is strongest when the probe tone is embedded in predictable sequences (those with low Sequence Predictive Information or high Sequence Coding Gain). To illustrate this interaction, we examine the influence of Probe Tone Information Content on Expectedness in stimuli with the lowest and highest Sequence Predictive Information. The correlation between Probe Tone Information Content and Expectedness is high for sequences in the lowest tercile of Sequence Prediction Information (values ranging from 0.04 to 0.16), $r(22) = -.98$, $p < .001$, but low for sequences in the upper tercile (values ranging from 0.51 to 0.77), $r(22) = -.32$, $p = .13$.

----------Insert Figure 3 about here----------

Probe Tone Information Content had the largest effect on Expectedness ratings.[4] For visualization of this effect in Fig. 3, the *average* expectedness rating for each melody is shown to display more clearly the relationship. In this analysis, Probe Tone Information Content shows a significant linear relationship with Expectedness Ratings, $R^2 = .69$, $F(70) = 154.20$, $p < .01$. Low Information Content tones receive consistently higher expectedness ratings than high Information Content probe tones over the course of listening.

2.3.2. Recognition memory in test sessions

Data from the test sessions are reported in Table 2 as *Proportion Correct Response*. Hits are defined as correctly identifying Old items as "heard before", and Correct Rejections are defined as responding that a New item has not been heard before. Chance performance is 0.5, and the similarity of performance for *Old* and *New* items indicates little bias towards either response.

----------Insert Table 2 about here----------

Despite poor recognition memory overall, we examined whether performance differed depending on the statistical properties of the individual sequences and whether this reflects learning of the statistical structure of the stimuli. To this end, a signal detection analysis was performed. Because the contrast of interest was the effect of different sequence properties (rather than differences between subjects), D-Prime and Criterion values were calculated for each sequence rather than by subject (for an example of this approach, see Dean, Harper, &

[4] Note that probe tones could occur between tones 17 and 23 of each tone sequence, and that in this regression analysis, whole sequence measures are computed for the entire sequence of tones. For comparison, the same regression analysis as the one described above was performed using sequence measures calculated *only to the point of the probe tone* in each sequence. Because the probe tone always occurred near the end of each stimulus, sequence measures reflecting the entire sequence were very highly correlated with measures based only on the events prior to and including the Probe tone. Therefore, and not surprisingly, this analysis yielded the same results as the regression analysis presented above. Using sequence measures computed for the entire sequence allows use of the same measures for Expectedness Ratings and Memory test results.

McAlpine, 2005). In addition, a Shapiro-Wilk normality test confirmed that the three information-theoretic measures were all non-normally distributed ($p < .05$), mandating non-parametric analysis. Therefore, Spearman rank correlations were calculated with Sequence Information Content, Sequence Coding Gain, and Sequence Predictive Information, respectively, as predictors for D-Prime and Criterion. There was a significant effect of each measure on D-Prime: *Sequence Information Content*, $r(22) = -.62$, $p < .01$; *Sequence Predictive Information*, $r(22) = -.51$, $p = .01$; and *Sequence Coding Gain*, $r(22) = .45$, $p < .05$. More predictable sequences (with low Information Content, low Predictive Information, or high Coding Gain) yielded higher D-Prime values (better discrimination between Old and New sequences) than unpredictable sequences (those with high Information Content, high Predictive Information or low Coding Gain).

Findings were similar for Criterion values, with each information-theoretic measure producing a significant effect: *Sequence Information Content*, $r(22) = .59$; $p < .01$, *Sequence Predictive Information*, $r(22) = .48$; $p < .05$, and *Sequence Coding Gain*, $r(22) = -.41$, $p < .05$. Sequences with low average Information Content and Predictive Information yielded lower Criterion values (more Hits) than those with high Information Content and Predictive Information, while sequences with low average Coding Gain yielded higher Criterion values than those with high Coding Gain.

2.3.2.1. Recognition memory test regression analysis

A logistic regression was performed to further assess the effects of the information-theoretic measures on recognition scores, and to explore whether these measures have a dynamic influence with increasing exposure to the tone sequences. As before, a stepwise regression was first performed to determine which factors to include in the logistic regression analysis, and the chosen factors were *Sequence Information Content*, *Sequence Coding Gain*,

Sequence Predictive Information, *Familiarity* (*Old* or *New*), and *Listening Session*, with Correct Response as the binary dependent variable.

All three sequence measures showed significant main effects: *Sequence Information Content* ($\chi^2(1) = 16.21$, $p < .001$), *Sequence Predictive Information* ($\chi^2(1) = 12.09$, $p < .001$), and *Sequence Coding Gain* ($\chi^2(1) = 4.27$, $p < .05$). *Listening Session* interacted significantly with each of the three sequence measures: *Sequence Information Content* X *Listening Session* ($\chi^2(2) = 6.14$, $p < .05$), *Sequence Predictive Information* X *Listening Session* ($\chi^2(2) = 7.98$, $p < .05$), and *Sequence Coding Gain* X *Listening Session* ($\chi^2(2) = 6.53$, $p < .05$). There was no significant main effect of Listening Session ($\chi^2(2) = 0.55$, $p = .76$). The impact of the three sequence measures all followed the same pattern: no impact on memory performance was evident initially, but by the third listening session, the measures were significantly correlated with Correct Response. Sequence Information Content and Sequence Predictive Information were negatively correlated with Correct Response, with higher values of these measures leading to fewer correct responses by the last listening session. Sequence Coding Gain was positively correlated with Correct Response, with lower values leading to fewer correct responses by the end of the study.

The only significant interaction including Familiarity was with Sequence Predictive Information ($\chi^2(1) = 12.15$, $p < .001$). As shown in Fig. 4, *New* sequences that are high in Predictive Information yield more correct responses than those with low Predictive Information. *Old* sequences show the opposite trend, with worse recognition memory performance on high Predictive Information sequences.

----------Insert Figure 4 about here----------

2.3.2.2. Confidence ratings

Confidence ratings for recognition memory judgments were collected after every test sequence; responses were made on a 1-5 scale where '1' represented *not confident* and '5'

represented *very confident*. A stepwise regression was used to select the factors to include in the ordinal logistic regression for confidence ratings, and as above, those selected were Sequence Information Content, Sequence Coding Gain, Sequence Predictive Information, Familiarity (Old or New stimulus), and Listening Session. This analysis yielded significant effects of *Sequence Information Content* ($\chi^2(1) = 11.87$, $p < .01$), *Sequence Predictive Information* ($\chi^2(1) = 12.50$, $p < .01$), and *Sequence Coding Gain* ($\chi^2(1) = 11.59$, $p < .01$), and the interaction of *Sequence Information Content* X *Listening Session* ($\chi^2(8) = 7.92$, $p < .05$).

As hypothesized, listeners made more confident memory judgments when sequences had lower Information Content, lower Predictive Information and higher Coding Gain. High Information Content sequences showed a decreasing level of confidence from session 1 to session 3.

2.4. Discussion

The results shed light on implicit statistical learning of novel sequential stimuli and go beyond previous research in showing, first, that statistical learning has an impact on recognition memory performance for tone sequences and, second, that expectations for individual tones, based on knowledge acquired through statistical learning, are dependent on the predictability of the entire stimulus. The results also highlight the information-theoretic properties that underlie these effects.

Regarding expectations for individual tones, there was a consistent negative relationship between Information Content and expectedness (such that high Information Content tones elicit greater unexpectedness) as hypothesized based on previous research with stylistically familiar stimuli (e.g., Hansen & Pearce, 2014; Pearce, 2005; Pearce et al., 2010). Surprisingly, there were no main or interaction effects of listening session, suggesting that, beyond the first session, expectations did not vary with increasing exposure to the stimuli. The primary novel finding is that this relationship between Information Content and

expectedness depends on the overall predictability of the preceding context, as hypothesized based on predictive coding theory (Clark, 2013; Friston, 2010; Friston & Kiebel, 2009). Specifically, while the negative correlation between probe Information Content and expectedness was observed for stimuli with low Sequence Predictive Information, there was no correlation for stimuli with high Sequence Predictive Information. The same pattern was also evident with Sequence Coding Gain, though the effect was weaker. The pattern was not evident for sequence Information Content but it is likely that this is because probe Information Content is highly correlated with sequence Information Content due to the way in which the stimuli and probe positions were generated and selected. The results suggest that unpredictable sequences (those with high sequence Predictive Information or low Coding Gain) compromise listeners' generation of strong expectations such that all probe tones (whether low or high in Information Content) are rated as moderately expected, which is indicative of an uncertain (or high entropy) prediction.

Regarding recognition memory, overall performance was consistently poor across the three sessions. In spite of this, performance for individual stimuli (measured using D-Prime and Criterion) did systematically vary with all three sequence measures. Unpredictable sequences (those with high Information Content, high Predictive Information, or low Coding Gain) produced worse memory performance and lower confidence ratings than predictable sequences (with low Information Content, low Predictive Information, or high Coding Gain). Interestingly, the influence of these sequence measures on recognition memory increased with exposure to the stimuli. Sequence Information Content did not have an impact on memory performance initially, but by the third session, was negatively correlated with Correct Response. This pattern was accompanied by decreasing confidence ratings over test sessions for high Information Content sequences. For memory performance (though not confidence), the same effect was found with Sequence Predictive Information and Coding

Gain. Furthermore, the significant interaction between Familiarity (Old vs. New stimulus) and Sequence Predictive Information is consistent with the hypothesis that unpredictability (due in this case to high Predictive Information) would impair initial encoding of familiar (Old) stimuli, making them more difficult to recognize subsequently, but would increase the distinctiveness of unfamiliar (New) test items, facilitating accurate discrimination and making correct rejection easier.

3. Computational simulations

In this section, we develop computational simulations of the behavioral data that shed light on the cognitive representations and processes involved in implicit statistical learning of this novel stimulus domain. The pattern of results reported above suggests a trajectory of changing memory performance, becoming increasingly associated with the information-theoretic properties of the stimuli over time. However, the generative and analytical models (used, respectively, to create the stimuli and compute their information-theoretic properties) only reflect learning within stimulus sequences and not across stimuli within the session as a whole.

To simulate statistical learning both across and within stimuli, we use a probabilistic computational model of auditory expectation (IDyOM, Information Dynamics of Music) developed, and described in detail, in previous research (Pearce, 2005).[5] IDyOM implements a model of statistical learning and has been shown to accurately account for listeners' pitch expectations in behavioral, physiological and EEG studies (e.g., Pearce, 2005; Pearce et al., 2010a; Omigie et al., 2012, 2013; Egermann et al., 2013; Hansen & Pearce, 2014), and simulate auditory boundary perception (Pearce, Müllensiefen, & Wiggins, 2010b). In many circumstances, IDyOM provides a more accurate model of listeners' pitch expectations than

[5] Software and documentation are available from https://code.soundsoftware.ac.uk/projects/idyom-project

static rule-based models (e.g., Narmour, 1990; Schellenberg, 1996, 1997), suggesting that expectation reflects a process of statistical learning and probabilistic generation of predictions (Hansen & Pearce, 2014; Pearce, 2005; Pearce et al., 2010a).

IDyOM has not yet been investigated as a cognitive model of memory for auditory sequences providing a further motivation for the present simulations. The following section gives a summary of the main features of IDyOM before presenting in detail the parameters varied in the simulations. Comparing models with different parameters against the results of the behavioral study allows inferences to be made about the cognitive mechanisms that underlie listeners' performance.

3.1. IDyOM model

IDyOM learns dynamically about sequential dependencies in the auditory environment to which it is exposed and generates probabilistic predictions about properties of events (pitch in the present case) for each tone in a tone sequence, given the context of the preceding sequence. The output is a conditional probability distribution predicting the pitch of the next tone, from which the estimated probability of the actual next tone may be extracted. Information content is the negative log probability of a tone (see Section 2) and reflects the unexpectedness of that tone from the perspective of the model. Previous research has shown that Information Content generated by IDyOM accurately simulates listeners' pitch expectations (Hansen & Pearce, 2014; Pearce, 2005; Pearce et al., 2010a; Omigie et al., 2012). In comparison to the first-order models used to generate and analyze stimuli for the behavioral study, IDyOM is a sophisticated variable-order Markov model (Begleiter et al., 2004) that has a flexible representation scheme (Conklin & Witten, 1995) and can combine information from short-term and long-term models. We now describe these features in further detail, to the extent that they bear on the present simulations.

IDyOM is based on a Markov or *n*-gram model (Manning & Schütze, 1999, ch. 9), which computes the conditional probability of a note given the $n - 1$ preceding notes in the melody. The quantity $n - 1$ is called the *order* of the model. Basic Markov models, such as the model used to generate the stimuli, have a fixed order. For the present stimuli, a zeroth-order model is simply the frequency of occurrence for each of the seven possible tones. A first-order model is a transition matrix containing the frequency with which each of the seven tones appears following each tone at the immediately preceding position in the sequence. Fixed-order models suffer from a variety of problems including the question of selecting the appropriate order and the so-called *zero-frequency* problem – how to estimate a non-zero probability for a tone that has not yet appeared in a particular context. IDyOM addresses these problems using methods developed in research on data compression (Bell, Cleary & Witten, 1990; Bunton, 1997) and statistical language modeling (Begleiter et al., 2004; Manning & Schütze, 1999). First, the order is allowed to vary at different points in the sequence and, second, a weighted average of probabilities is computed from models of different order, a process known as smoothing (Begleiter et al., 2004; Bell, Cleary & Witten, 1990; Bunton, 1997; Manning & Schütze, 1999). The maximum order may be fixed at a particular value or may be free to vary, in which case the longest matching context is used, which may vary at each position in a sequence (Bunton, 1997). In the present research, we compare variable-order models with models whose order is limited to zero (zeroth-order models) and one (first-order models).

IDyOM has two subcomponents that may be configured to be used either individually or together. The first is a Long-Term Model (LTM) that is trained on an entire corpus (simulating learning based on a listener's long-term schematic exposure to music), and the second is a Short-Term Model (STM), which is exposed incrementally to the current stimulus (simulating local learning of structure and statistics in the current listening episode). The

LTM is static once trained while a variant of this subcomponent, called the LTM+, continues to learn dynamically from the sequences to which it is subsequently exposed. The LTM+ may or may not be pre-trained on a corpus like the LTM. While the LTM+ learns incrementally across stimuli, the STM begins each new stimulus as a *tabula rasa* and learns incrementally within that tone sequence without carrying any learning over between stimuli. The distributions generated by the STM and LTM/LTM+ may be combined – various approaches are possible but here we use a geometric mean, weighted by the entropy of the distribution generated by each subcomponent (Conklin & Witten, 1995; Hinton, 2002; Pearce, 2005). Combining the STM and LTM yields a BOTH configuration while combining the STM and LTM+ yields a BOTH+ configuration.

The LTM+ corresponds to the cognitive model assumed in many studies of statistical learning (e.g., Saffran et al., 1999), although learning is often assumed to take place only for the exposure stimuli and not for test items (which *are* included in the LTM+). Given pervasive evidence of dynamic learning across stimuli, including test items (e.g., Rohrmeier et al., 2011), the STM and LTM variants are not examined in the present research, leaving the LTM+ and BOTH+ configurations for simulations of listeners' responses. The LTM+ and BOTH+ models used in the present simulations have no prior training before being exposed to the stimuli making up an experimental session. The analytical model introduced in Section 2 does not correspond to any of the IDyOM configurations, but could be described as an STM initialized with a bigram table estimated from a corpus of Western tonal music.

Finally, IDyOM has the ability to use different pitch features (e.g., chromatic pitch, sequential pitch interval) to predict properties of tones (see Conklin & Witten, 1995; Pearce, 2005) which is important given evidence that listeners represent pitch in different ways (e.g., Levitin & Tirovolas, 2009; Shepard, 1982). The present research focuses on models that use either pitch or pitch interval representations.

3.2. Method

The models selected for the simulation varied in terms of three factors: model configuration (LTM+ or BOTH+), model order (zero, first, or variable order), and feature (chromatic pitch or pitch interval). The output of the resulting 12 models was examined to find which best simulates the mean Expectedness ratings and D-Prime memory scores from the behavioral study. To compare the IDyOM simulations to listeners' responses we use the information content of the specified probe tone in the case of expectations. For simulating memory performance, we use the average information content for the whole stimulus, under the hypothesis that more unpredictable stimuli should be less accurately encoded (see Section 2.1.2). We complement these analyses of model fit with human expectedness and memory performance with additional analyses of intrinsic model performance in predicting the stimuli. We use information content as a measure of prediction performance because low information content indicates that the model is able to predict a stimulus with high probability. These analyses are conducted both for the information content of probe tones in the listening sessions and for mean information content of each stimulus in the test sessions.

3.3. Hypotheses

First and foremost, we hypothesize that IDyOM will be able to successfully simulate listeners' performance on both the expectation and memory tasks from the behavioral study and, specifically, that expectedness for individual tones will correlate negatively with information content, while memory performance will correlate negatively with the information content averaged across all notes in a sequence.

The BOTH+ models are expected to simulate listeners' performance better than LTM+ models because they are more cognitively plausible, simulating both local learning within a stimulus (the STM) and long-term incremental learning of statistical structure with

increasing exposure to the stimuli (the LTM+). This configuration reflects a long history of research on human memory that has incorporated both short-term and long-term components (Baddeley, Papagno, & Vallar, 1988; Hulme, Maughan, & Brown, 1991; Ericsson & Kintsch, 1995).

The chromatic pitch feature and pitch interval feature are compared to ascertain whether listeners represent the stimuli in terms of absolute or relative pitch structure. We hypothesize that the pitch feature will be optimal for modeling pitch *expectation* (as we specifically requested listeners to rate the expectedness of the pitch of the probe tone), while pitch interval information will best simulate *memory* for sequences, because pitch interval structure has been shown to be important in memory for melodic sequences (e.g., Dowling, 1991).

Finally, we compare models with a fixed order-bound of zero, a fixed order bound of one, and a variable order bound (no fixed order), to examine whether listeners are taking advantage of any higher-order structure in the stimuli. Listeners are sensitive to zeroth-order pitch distributions in music (Krumhansl, 1990; Oram & Cuddy, 1995) but also show influence of higher-order statistical structure on their expectations (Hansen & Pearce, 2014; Krumhansl et al., 2000). In the present study, because the stimuli were generated using a first-order pitch model, we postulate that first-order pitch models or zeroth-order interval models (since an interval spans two pitches, a zeroth-order interval model is more comparable to a first-order pitch model than a first-order interval model), would best simulate listeners' responses, with only limited benefit from using higher-order models.

3.4. Results

The results of the 12-model comparison for expectedness ratings from the behavioral study are shown in Table 3. A BOTH+ zeroth-order model with an interval feature best simulated listeners' expectedness responses ($r(22) = -.86$, $p < .01$), providing the highest

correlations in each listening session and overall. No statistical difference, however, was found between this and the next best performing model, a BOTH+ first-order model also using the interval feature (Williams' $t(69) = 1.42$, $p = .16$). The model did perform better than the third-best model, a BOTH+ zeroth-order model using the chromatic pitch feature (Williams' $t(69) = 1.97$, $p = .05$).

----------Insert Table 3 about here----------

The results of the 12-model comparison for memory test D-Prime scores from the behavioral study are shown in Table 4. Parsimoniously, the same BOTH+ model configuration using the pitch interval feature yields the highest correlation with average memory performance ($r(22) = -.72$, $p < 0.01$). As for the expectedness results, no statistically significant difference was found between the zeroth and first-order models for this BOTH+ interval configuration (Williams' $t(21) = 0.76$, $p = .45$), and again the best-fitting model's performance was superior to the third-best performing model (an LTM+ first order model using pitch interval, Williams' $t(21) = 2.36$, $p < .05$). Interestingly, all models performed poorly in the first test session, but correlations increased for later test sessions (see Table 4), suggesting that listeners' memory performance was increasingly well simulated by the models with each listening session.

----------Insert Table 4 about here----------

Turning now to model performance, probe information content generated by the BOTH+ pitch-interval models in the three listening sessions was submitted to a 3x3 ANOVA with *model order* (0, 1, variable) and *session number* (1, 2, 3) as independent variables. The results show a significant main effect of model order, $F(2, 207) = 8.28$, $p < .01$, but no main effect or interaction effect of session number. Post-hoc t-tests suggest that the significant main effect of model order arises because the information content of the variable-order model is significantly different from that of both the first-order model, $t(44.98) = 2.28$, $p = .03$, and

the zeroth-order model, $t(38.9) = 2.15$, $p = .04$, while the latter two models do not differ significantly, $t(42.3) = 0.27$, $p = .79$. Finally, average sequence information content generated by the BOTH+ pitch-interval model in the three test sessions was submitted to a 3x3 ANOVA with *model order* (0, 1, variable) and *session number* (1, 2, 3) as independent variables. The results showed no significant main or interaction effects.

3.5. Discussion

As hypothesized, the information content of probe tones returned by IDyOM accurately simulated expectedness ratings, accounting for up to 74% of the variance overall. Information content did not vary with increasing exposure, as with listeners' expectations. Of particular interest, however, is the novel finding that overall memory performance showed significant correlations with mean information content, suggesting that memory is impaired for more unpredictable sequences (those with higher information content). Further, the results corroborate the finding that listeners' memory performance dynamically changes across test sessions, with performance becoming more strongly aligned to the information content of sequences over the course of exposure. Information content accounts for up to 85% of variance in memory performance by the final session.

In both cases, parsimoniously, the best-fitting models are zeroth- and first-order BOTH+ pitch-interval models (the fit of zeroth- and first-order models is statistically indistinguishable). The BOTH+ configuration indicates that dynamic statistical learning takes place both within individual stimuli (simulated by the STM component) and across stimuli throughout the course of the experimental session (simulated by the LTM+ component).[6] The

[6] The fact that the BOTH+ provides a better fit than the LTM+ demonstrates that the STM component makes a significant contribution to simulating listeners' responses. The BOTH+ interval model also produced higher correlations with both memory performance and expectedness ratings than first- and zeroth-order STM interval models, demonstrating the contribution of the LTM+. Finally, the best-performing IDyOM model also yields higher

fact that pitch interval representations best match human performance is consistent with our hypothesis for memory performance, given strong prior evidence (Dowling, 1991). However, it is inconsistent with our hypothesis for pitch expectations and also inconsistent with the models used to generate and analyze the stimuli in Section 2. Pitch interval is a more abstract representation than pitch, since an interval can map onto many concrete pairs of pitches. It seems likely that this more abstract representation facilitates better generalization of statistical learning across stimuli, both for IDyOM and for listeners.

Given that the stimuli were generated using a first-order model combined with ample evidence that listeners are capable of learning first-order statistics (e.g., Romberg & Saffran, 2010), it might seem surprising that the first-order models do not account better than zeroth-order models for both expectedness and memory performance. Recall, however, that the input to the best-fitting models is the pitch interval (i.e., log frequency difference) between successive tones in the stimuli. Therefore, a zeroth-order pitch-interval model actually represents statistical information about pairs of tones. In this way, it is similar to the first-order pitch model that was used to generate the stimuli and, therefore, may encode first-order statistical regularities in pitch. Furthermore, the IDyOM performance results show that there is not significantly more first-order pitch-interval information in the stimuli than zeroth-order pitch-interval information, so the lack of difference between first- and zeroth-order models in fit with human performance is consistent with the statistical structure of the stimulus.

The fact that the variable-order IDyOM model produced lower mean information content for the probe tones than the zeroth- and first-order models, suggests the presence of higher-order structure in the stimuli (i.e., better prediction from contexts longer than one tone). This may reflect IDyOM recognizing exact repetitions of stimuli, but the effect was not replicated for the average information content of stimuli in the test sessions. Certainly, there

---

correlations with expectedness ratings and memory performance than the analytical model introduced in Section 2.

was no evidence that participants made use of any higher-order structure and the results for memory performance also suggest that they were learning generalized statistical structure rather than individual stimuli.

Regarding the effects of exposure, neither probe information content, nor expectedness, nor the correlation between the two vary across sessions, suggesting that both listeners and IDyOM learn pointwise statistical regularities present in the stimuli relatively quickly during the first session. Interestingly, however, the correlations between memory performance and information content do increase dramatically over the course of the three memory test sessions. As with probe information content, there is no evidence that mean sequence information content varies across the three sessions, suggesting again that IDyOM learns the regularities present in the stimuli relatively quickly during the first session. However, listeners' memory performance becomes more systematically related to information content with increasing exposure. Mean information content (from the zeroth- and first-order pitch interval BOTH+ models respectively) accounts for 16% and 23% of the variance in memory performance for Session 1. This rises to 66% (for both models) in Session 2, and rises again to 85% and 83% in Session 3. While information content provides a convincing account of memory performance in the final session, the IDyOM simulations do not explain why memory performance shows weaker correlations with information content in the first two sessions. Explanations for this trajectory are considered further in Section 4.

Finally, because the stimuli used pitches drawn from a diatonic scale, we investigated the performance of IDyOM models for which the LTM+ was given pre-training on a corpus of Western tonal music. Pre-training did not improve performance for either the Expectedness or Memory simulations, suggesting that listeners dynamically and implicitly acquired a new cognitive representation of statistical regularities in the stimuli over the

course of the study. For full details and results of these simulations please consult the Appendix.

4. General Discussion

Whereas existing research on auditory statistical learning has focused either on segmentation (e.g., Romberg & Saffran, 2010) or artificial grammar learning (e.g., Perruchet & Pacton, 2006; Rohrmeier et al., 2011), the present research examines the effects of auditory statistical learning on expectation and recognition memory, using a combination of behavioral investigation and computational simulation. The results suggest that performance is significantly related to information-theoretic properties of the stimuli and that the cognitive processes involved can be simulated using low-order probabilistic models that dynamically learn the statistical regularities in pitch interval both within and across stimuli. Of particular interest is the novel finding that recognition memory performance is systematically related to the information-theoretic predictability of stimuli. We discuss these findings in terms of the literature on auditory perception, statistical learning and predictive coding.

4.1 Expectedness

The results confirm the relationship between auditory expectations and information content arising from statistical learning of these sequences. Importantly, previous research demonstrating this relationship has used music as a domain in which listeners already have prior experience (Hansen & Pearce, 2014; Pearce, 2005; Pearce et al., 2010a; Omigie et al., 2012). The present results extend these findings to relatively short-term learning of a novel set of auditory stimuli, providing further evidence for IDyOM as a model of domain-general predictive mechanisms in cognitive processing of auditory sequences.

Following exposure to a novel set of artificially-constructed auditory sequences, listeners generate expectations that reflect the models used to generate the individual

sequences (see also Loui et al., 2010). However, their expectations are better characterized by models that dynamically learn the statistics of pitch interval relationships within and across stimuli. Statistical learning of pitch intervals has been previously demonstrated (Saffran & Griepentrog, 2001; Saffran et al. 2005) and we have argued that in the present context, pitch intervals afford a more abstract representation than pitch, allowing more powerful generalization of statistical patterns across stimuli.

The results also show that expectations for events depend not only on the information content of the current event but also on the overall predictability of the context within which it is embedded. In predictable contexts, listeners generate expectations that conform to the probabilistic model but for tones embedded in unpredictable contexts, they generate moderate expectations, regardless of the information content of the tone itself. Predictive coding theory suggests that to maintain an accurate model of the sensory signal, an agent must modulate top-down predictions to minimize surprise (Clark, 2013; Friston, 2010; Friston & Kiebel, 2009). The present results are consistent with this, suggesting that unpredictable sequences do not allow for the generation of strong, specific top-down predictions but instead produce uncertain, high-entropy expectations even when there is predictable structure in the signal.

The IDyOM models that best account for listeners' expectation (and memory) performance integrate information generated by dynamic statistical learning of two kinds: first, information learned incrementally within each stimulus; second, information learned incrementally across all stimuli in the study. This is congruent with evidence that listeners are sensitive to statistical information in auditory sequences on both short (e.g., Oram & Cuddy, 1995) and long timescales (e.g., Krumhansl, 1990). The relative influence of the top-down predictions from IDyOM's long- and short-term models is adjusted according to the entropy, or uncertainty, of the distributions they generate. Once again, this approach is compatible

with predictive coding theory, where prediction errors are weighted (through synaptic gain) by their precision or uncertainty (Friston, 2010).

Both the dynamic nature of the learning and the effects of short-term learning within stimuli have implications for existing research on implicit statistical learning of unfamiliar sequential stimuli (e.g., Conway & Christiansen, 2005; Loui et al., 2010; Perruchet & Pacton, 2006; Saffran, 2003b). In particular, research on implicit statistical learning must now assume that participants learn throughout an experimental session, both within and across stimuli – learning does not end with the exposure phase (see also Rohrmeier et al., 2011). In future research, it would be useful to explicitly examine learning within and across stimuli by systematically varying the degree of statistical structure that is shared across individual stimuli.

4.2 Recognition Memory

Overall memory performance was consistently poor across the three sessions, in line with previous research using more stylistically familiar stimuli (e.g., Halpern & Bartlett, 2010; Dowling, Bartlett, Halpern, & Andrews, 2008; Halpern and Müllensiefen, 2008; Bartlett, Halpern, & Dowling, 1995). Although the overall proportion of correct responses remained similar across the three test sessions, the types of errors listeners made varied as a function of the statistical properties of the sequences and the degree to which the listener had been exposed to them. This result confirms the hypothesized relationship between information-theoretic measures of predictability and recognition memory, and provides the beginnings of a plausible computational model of the cognitive processing underlying findings that stylistically unfamiliar or complex stimuli yield poor recognition memory (Cuddy et al., 1981; Halpern & Bartlett, 2010; Näätänen, Schröger, Karakas, Tervaniemi, & Paavilainen, 1993).

What underlies these effects? We hypothesized that unpredictable sequences would be difficult to encode, leading to poorer memory performance for familiar stimuli when they are presented again, but also that unfamiliar unpredictable sequences presented as foils in the test phases might be more distinctive, thereby facilitating accurate discrimination. The results validate these hypotheses because unpredictable stimuli (specifically, those with high predictive information) produce more misses for targets (familiar test items) but also more correct rejections of foils. The present results provide evidence regarding the information-theoretic stimulus properties and underlying statistical model that account for the perception of predictability and distinctiveness in causing these effects. Further research is required to examine whether this pattern of results extends to stylistically familiar materials and other domains, such as language.

Previous research on statistical learning using segmentation tasks (Creel et al., 2004; Saffran et al., 1996a,b; Saffran et al., 1999; Saffran et al., 2005; Saffran & Griepentrog, 2001) demonstrates that individuals can learn the statistical structure of artificial sequences and use these learned representations to segment where first-order probabilities are low (see also Brent, 1999; Elman 1990; Pearce et al., 2010b). The present research extends these findings to the effects of statistical learning on the cognitive processes of expectation and memory. The segmentation task depends on these processes because participants must generate probabilistic expectations based on the learned statistical properties of the exposure phase in order to identify grouping boundaries where transition probabilities are low. They must also hold sequences in memory in order to match incoming sequences to the learned model (though the sequences are much shorter than those used in the present research). Conversely, it is possible that the memory advantages observed in the present research for predictable stimuli (of much greater length) arise because these stimuli can be represented in memory as a smaller number of longer chunks than is the case for unpredictable stimuli.

With repeated listening, memory performance became significantly more aligned with the information-theoretic predictability of the stimulus sequences. Overall memory performance did not improve over the three listening sessions but, with increasing exposure, performance became worse, and confidence became lower, for unpredictable stimuli. This suggests that participants were learning the statistical regularities describing the stimuli rather than the particular exemplars themselves (Cleeremans et al., 1998; Halpern & Bartlet, 2010; Loui et al., 2010; Saffran et al., 1999; Stadler & Frensch, 1998). Had we used fewer and/or more brief stimuli, recognition memory performance might have been better (see Loui & Wessel, 2008), but this would also likely have resulted in less abstraction of the underlying statistical relationships across stimuli (Cleeremans et al., 1998).

Expectations are well simulated by information content in the first session and do not change thereafter, suggesting that statistical learning occurs primarily within the first session. However, recognition memory shows a trajectory across sessions, with information content accounting increasingly well for performance from Session 1 (23% of variance explained) to Session 3 (85% of variance explained). Thus while IDyOM simulates memory performance very well in the final session, it does not account for this trajectory. We consider two possible interpretations. First, it may be that statistical learning takes place across all three sessions, influencing memory performance, but stops having an impact on expectation after the first session (for example, due to ceiling effects if the expectation task is much easier than the memory task). Second, it may be that statistical learning takes place primarily within the first session but that further exposure is required for the listener's learned internal predictive model to have an impact on the encoding and consolidation of memory representations for the stimuli. The two interpretations are not mutually exclusive and further research is required to disentangle them.

Overall the simulation results suggest that learning of statistical structure is rapid, continuous and implicit, and takes place dynamically and continuously both within and across stimuli, incrementally developing more accurate cognitive models of the statistical structure of the stimuli. Importantly, the results suggest that when confronted with an unfamiliar auditory environment, rather than updating an existing cognitive representation, listeners construct a new cognitive model to describe the statistical structure of the environment. This is consistent with a central tenet of predictive coding theory (Clark, 2013; Friston, 2010), that perception is a process of active inference and continuous refinement to achieve parsimonious models of the sensory environment (see also Barlow, 1959; Dayan et al., 1995; Gregory, 1980; Helmholtz, 1866). It would be interesting to examine how much and what kinds of unpredictability can be tolerated before prior models are discarded. This question has implications for research on auditory statistical learning more widely where artificially constructed auditory sequences are in some cases assumed to invoke a prior model (e.g., Oram & Cuddy, 1995) and in other cases not (e.g., Loui et al., 2010; Saffran et al., 1999).

4.3. Conclusion and future directions

In summary, the behavioral study and IDyOM simulations clarify how listeners represent and process novel sequential auditory stimuli. The results indicate that information-theoretic properties of sequential stimuli have an impact on both expectation and recognition memory. Furthermore, they provide the beginnings of a quantitative model of the dynamic cognitive processes involved in implicit statistical learning of the regularities present in novel sequential auditory stimuli. In the process of acquiring an internal predictive model, statistical information appears to contribute *over different time-scales* to ongoing sequential processing, consistent with hierarchical predictive coding theory (Clark, 2013): first, momentary expectations are based on the information content of the current event; second, the

predictability of the entire stimulus has an impact on expectation and memory processing; and third, the properties of the entire stimulus set, learned incrementally over time, influence expectation and memory for tone sequences. Further research might extend the present results to stimuli containing temporal structure, higher-order statistical relationships (Hansen & Pearce, 2014; Krumhansl et al., 2000) and relationships between non-adjacent events in pitch sequences (Creel et al., 2004).

References

Agres, K., Abdallah, S., & Pearce, M. (2013). An Information-Theoretic Account of Musical Expectation and Memory. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 127–132). Austin, Texas: Cognitive Science Society.

Abdallah, S. & Plumbley, M. (2009). Information dynamics: patterns of expectation and surprise in the perception of music. *Connection Science, 21,* 89-117.

Abdallah, S. & Plumbley, M. (2010). Predictive information, Multi-information, and Binding information. Technical Report C4DM-TR10-10, Centre for Digital Music, Queen Mary University of London.

Abdallah, S. & Plumbley, M. (2012). A measure of statistical complexity based on predictive information with application to finite spin systems. *Physics Letters,* 376, 275-281.

Ames, C. (1989). The Markov process as a compositional model: A survey and tutorial. *Leonardo*, 22, 175–187.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. *The Psychology of Learning and Motivation*, *2*, 89-195.

Baddeley, A., Papagno, C., & Vallar, G. (1988). When long-term learning depends on short-term storage. *Journal of memory and language*, *27*(5), 586-595.

Bar, M. (2007). The proactive brain: using analogies and associations to generate predictions. *Trends in Cognitive Sciences*, *11*, 280-289.

Barlow, H. B. (1959). Sensory mechanisms, the reduction of redundancy, and intelligence. *The Mechanisation of Thought Processes*, *10*, 535-539.

Bartlett, F. (1932). *Remembering: A study in experimental and social psychology.* Cambridge, UK: Cambridge University Press.

Bartlett, J. C., Halpern, A., & Dowling, W. J. (1995). Recognition of familiar and unfamiliar melodies in normal aging and Alzheimer's disease. *Memory & Cognition, 23*, 531-546.

Begleiter, R., El-Yaniv, R., & Yona, G. (2004). On prediction using variable order Markov models. *Journal of Artificial Intelligence Research*, *22*, 385–421.

Bell, T. C., Cleary, J. G., & Witten, I. H. (1990). *Text Compression.* Englewood Cliffs, NJ: Prentice Hall.

Bertuccelli, L., & How, J. (2008). Estimation of Non-stationary Markov Chain Transition Models. *Proceedings of the 47th IEEE Conference on Decision and Control*. Paper presented at The 47th IEEE Conference on Decision and Control, Cancun, Mexico, 9-11 December (pp. 55-60).

Bharucha, J. J. (1987). Music cognition and perceptual facilitation: A connectionist framework. *Music Perception*, 5, 1–30.

Bubic, A., Von Cramon, D. Y., & Schubotz, R. I. (2010). Prediction, cognition and the brain. *Frontiers in Human Neuroscience*, 4, 25. doi: 10.3389/fnhum.2010.00025.

Brent, M. R. (1999). Speech segmentation and word discovery: A computational perspective. *Trends in Cognitive Sciences, 3*, 294-301.

Bunton, S. (1997). Semantically motivated improvements for PPM variants. *The Computer Journal*, 40 (2/3), 76–93.

Carlsen, J. C. (1981). Some factors which influence melodic expectancy. *Psychomusicology*, *1*, 12-29.

Castellano, M. A., Bharucha, J. J., & Krumhansl, C. L. (1984). Tonal hierarchies in the music of north India. *Journal of Experimental Psychology: General*, *113*, 394-412.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*, 181-204.

Cleeremans, A., Destrebecqz, A., & Boyer, M. (1998). Implicit learning: news from the front. *Trends in Cognitive Sciences*, 2, 406-416.

Cohen, J. E. (1962). Information theory and music. *Behavioral Science*, 7, 137–163.

Conklin, D. & Witten, I. H. (1995). Multiple viewpoint systems for music prediction. *Journal of New Music Research*, 24, 51–73.

Conway, C. M., & Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 24-39.

Conway, C. M., & Christiansen, M. H. (2006). Statistical learning within and between modalities pitting abstract against stimulus-specific representations. *Psychological Science*, *17*, 905-912.

Creel, S. C., Newport, E. L., & Aslin, R. N. (2004). Distant melodies: Statistical learning of nonadjacent dependencies in tone sequences. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 30, 1119–1130.

Cristià, A., McGuire, G. L., Seidl, A., & Francis, A. L. (2011). Effects of the distribution of acoustic cues on infants' perception of sibilants. *Journal of Phonetics*, *39*(3), 388-402.

Cuddy, L., Cohen, A., & Mewhort, D. (1981). Perception of structure in short melodic sequences. *Journal of Experimental Psychology: Human Perception and Performance, 7*, 869-883.

Dayan, P., Hinton, G. E., Neal, R. M., & Zemel, R. S. (1995). The Helmholtz machine. *Neural Computation*, *7*, 889-904.

Dean, I., Harper, N., & McAlpine, D. (2005). Neural population coding of sound level adapts to stimulus statistics. *Nature Neuroscience, 8*, 1684-1689.

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, *8*, 1117-1121.

Dienes, Z., & Longuet-Higgins, C. (2004). Can musical transformations be implicitly learned? *Cognitive Science*, *28*, 531-558.

Dowling, W. J., (1991). Tonal strength and melody recognition after long and short delays. *Perception & Psychophysics, 50*, 305-313.

Dowling, W. J., Bartlett, J. C., Halpern, A., & Andrews, M. (2008). Melody recognition at fast and slow tempos: Effects of age, experience, and familiarity. *Perception & Psychophysics, 70*, 496-502.

Eerola, T. (2004). Data-driven influences on melodic expectancy: Continuations in North Sami Yoiks rated by South African traditional healers. In S. D. Lipscomb, R. Ashley, R. O. Gjerdingen, & P. Webster (Eds.), *Proceedings of the Eighth International Conference of Music Perception and Cognition* (pp. 83–87). Adelaide, Australia: Causal Productions.

Egner, T., Monti, J. M., & Summerfield, C. (2010). Expectation and surprise determine neural population responses in the ventral visual stream. *The Journal of Neuroscience*, *30*, 16601-16608.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science, 14*, 179-211.

Egermann, H., Pearce, M., Wiggins, G., McAdams, S. (2013). Probabilistic models of expectation violation predict psychophysiological emotional responses to live concert music. *Cognitive, Affective, & Behavioral Neuroscience. 13*, 533–553.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological review*, *102*, 211-245.

Fiser, J., & Aslin, R. N. (2002). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 458.

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, *11*, 127-138.

Friston, K., & Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*, 1211-1221.

Furl N., Kumar S., Alter K., Durrant S., Shawe-Taylor J., Griffiths T. D. (2011). Neural prediction of higher-order auditory sequence statistics. *NeuroImage*, **54**(3), 2267-2277.

Gregory, R. L. (1980). Perceptions as hypotheses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *290*, 181-197.

Hale, J. (2006). Uncertainty about the rest of the sentence. *Cognitive Science, 30*, 643-672.

Halpern, A., & Bartlett, J. (2010). Memory for melodies. In M.R. Jones, A. Popper, & R. Fay (Eds.), *Music Perception.* New York: Springer-Verlag.

Halpern, A. & Müllensiefen, D. (2008). Effects of timbre and tempo change on memory for music. *The Quarterly Journal of Experimental Psychology, 61*, 1371-1384.

Hansen, N., & Pearce, M. (2014). Predictive Uncertainty in Auditory Sequence Processing. *Frontiers in Psychology*, *5*. doi: 10.3389/fpsyg.2014.01052.

Helmholtz, H. V. (1866). Concerning the perceptions in general. In *Treatise on Physiological Optics* (Vol. III, pp. 1-37). (Translated by J. P. C. Southall, 1925, Optical Society of America. Reprinted in 1962, New York: Dover Publications, Inc.)

Hinton, G. (2002). Training Products of Experts by Minimizing Contrastive Divergence. *Neural Computation, 14*, 1771-1800.

Hulme, C., Maughan, S., & Brown, G. D. (1991). Memory for familiar and unfamiliar words: Evidence for a long-term memory contribution to short-term memory span. *Journal of Memory and Language*, *30*, 685-701.

Hunt, R. H., & Aslin, R. N. (2001). Statistical learning in a serial reaction time task: access to separable statistical cues by individual learners. *Journal of Experimental Psychology: General*, *130*, 658.

Huron, D. (2006). *Sweet anticipation: Music and the psychology of expectation*. Cambridge, MA: MIT Press.

Kessler, E. J., Hansen, C., & Shepard, R. N. (1984). Tonal schemata in the perception of music in Bali and in the West. *Music Perception*, 2, 131-165.

Kirkham, N. Z., Slemmer, J. A., and Johnson, S. P. (2002). Visual statistical learning in infancy: evidence for a domain general learning mechanism. *Cognition* 83, B35–B42. doi: 10.1016/S0010-0277(02)00004-5.

Knopoff, L. & Hutchinson, W. (1981). Information theory for musical continua. *Journal of Music Theory,* 25, 17–44.

Knopoff, L. & Hutchinson, W. (1983). Entropy as a measure of style: The influence of sample length. *Journal of Music Theory*, 27, 75–97.

Krumhansl, C. (1990). *Cognitive foundations of musical pitch*. Oxford, UK: Oxford University Press.

Krumhansl, C. & Kessler, E. (1982). Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychological Review, 89*, 334-368.

Krumhansl, C., Louhivuori, J., Toiviainen, P., Järvinen, T., & Eerola, T. (1999). Melodic expectation in Finnish spiritual folk hymns: Convergence of statistical, behavioral, and computational approaches. *Music Perception*, 17, 151-195.

Krumhansl, C., Toivanen, P., Eerola, T., Toiviainen, P., Järvinen, T., & Louhivuori, J. (2000). Cross-cultural music cognition: Cognitive methodology applied to North Sami yoiks. *Cognition, 76*, 13–58.

Kumar, S., Sedley, W., Nourski, K., Kawasaki, H., Oya, H., Patterson, R., Howard, III, R., Friston, K., & Griffiths, T. (2011). Predictive coding and pitch processing in the auditory cortex. *Journal of Cognitive Neuroscience*, 23, 3084-3094.

Levitin, D., & Tirovolas, A. (2009). Current Advances in the Cognitive Neuroscience of Music. *Annals of the New York Academy of Sciences,* 1156, 211–231.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition,* 106, 1126-1177.

Loui, P., & Wessel, D. (2008). Learning and linking an artificial musical system: Effects of set size and repeated exposure. *Musicae Scientiae,* 12, 207–230.

Loui, P., Wessel, D. L., & Kam, C. (2010). Humans rapidly learn grammatical structure in a new musical scale. *Music Perception, 27*, 377–388.

Manning, C. & Schütze, H. (1999). *Foundations of statistical natural language processing*. Cambridge, MA: MIT Press.

Meyer, L. (1956). *Emotion and Meaning in Music*. Chicago, IL: University of Chicago Press.

Mirman, D., Magnuson, J. S., Graf Estes, K. & Dixon, J. A. (2008). The link between statistical segmentation and word learning in adults. *Cognition,* 108, 271-280.

Moles, A. (1966). *Information Theory and Esthetic Perception*. Champaign, IL: University of Illinois Press.

Müllensiefen, D., & Halpern, A. (2014). The role of features and context in recognition of novel melodies. *Music Perception,* 31, 418-435.

Näätänen, R., Schröger, E., Karakas, S., Tervaniemi, M., & Paavilainen, P. (1993). Development of a memory trace for a complex sound in the human brain. *NeuroReport*, *4*, 503-506.

Narmour, E. (1990). *The analysis and cognition of basic melodic structures: The implication-realisation model.* Chicago: University of Chicago Press.

Omigie, D., Pearce, M. T., and Stewart, L. (2012). Tracking of pitch probabilities in congenital amusia. *Neuropsychologia*, 50, 1483-1493.

Omigie, D., Pearce, M. T., Williamson, V., & Stewart, L. (2013). Electrophysiological correlates of melodic processing in congenital amusia. *Neuropsychologia*, 51, 1749-1762.

Oram, N. & Cuddy, L. L. (1995). Responsiveness of Western adults to pitch-distributional information in melodic sequences. *Psychological Research*, 57, 103–118.

Pearce, M. T. (2005). *The construction and evaluation of statistical models of melodic structure in music perception and composition*. Doctoral Dissertation, Department of Computing, City University, London, UK.

Pearce, M. T. & Wiggins, G. A. (2012). Auditory expectation: The information dynamics of music perception and cognition. *TopiCS in Cognitive Science*, 4, 625-652.

Pearce, M. & Wiggins, G. (2006). Expectation in melody: The influence of context and learning. *Music Perception, 23,* 377–405.

Pearce, M., Ruiz, M., Kapasi, S., Wiggins, G., & Bhattacharya, J. (2010a). Unsupervised statistical learning underpins computational, behavioural and neural manifestations of musical expectation. *NeuroImage, 50,* 302-313.

Pearce, M., Müllensiefen, D., & Wiggins, G. (2010b). The role of expectation and probabilistic learning in auditory boundary perception: A model comparison. *Perception, 9,* 1367-1391.

Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Sciences*, *10*, 233-238.

Redington, M., Chater, N., & Finch S. (1998). Distributional information: A powerful cue for acquiring syntactic categories. *Cognitive Science,* 22, 425-469.

Rohrmeier, M., & Rebuschat, P. (2012). Implicit learning and acquisition of music. *Topics in Cognitive Science, 4*, 525–553.

Rohrmeier, M., Rebuschat, P. & Cross, I. (2011). Incidental and online learning of melodic structure. *Consciousness and Cognition*, 20, 214-222.

Romberg, A. R., & Saffran, J. R. (2010). Statistical learning and language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1, 906-914.

Saffran, J. R. (2003a). Statistical language learning mechanisms and constraints. *Current Directions in Psychological Science*, *12*, 110-114.

Saffran, J. R. (2003b). Musical learning and language development. *Annals of the New York Academy of Sciences*, *999*, 397-401.

Saffran, J. R., & Griepentrog, G. J. (2001). Absolute pitch in infant auditory learning: evidence for developmental reorganization. *Developmental Psychology*, *37*, 74.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996a). Statistical learning by 8-month-old infants. *Science*, 274, 1926-1928.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996b). Word segmentation: The role of distributional cues. *Journal of Memory and Language,* 35, 606-621.

Saffran, J.R., Johnson, E.K., Aslin, R.N., & Newport, E.L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition, 70*, 27–52.

Saffran, J. R., Pollak, S. D., Seibel, R. L., & Shkolnik, A. (2007). Dog is a dog is a dog: Infant rule learning is not specific to language. *Cognition*, *105*, 669-680.

Saffran, J. R., Reeck, K., Niebuhr, A., & Wilson, D. (2005). Changing the tune: the structure of the input affects infants' use of absolute and relative pitch. *Developmental Science*, *8*, 1-7.

Schellenberg, E. G. (1997). Simplifying the implication-realisation model of melodic expectancy. *Music Perception*, 14, 295–318.

Schellenberg, E. G. (1996). Expectancy in melody: Tests of the implication-realisation model. *Cognition*, 58, 75–125.

Shepard, R. (1982). Geometrical approximations to the structure of musical pitch. *Psychological Review, 89*, 305-333.

Stadler, M., & Frensch, P. (1998). *Handbook of implicit learning*. Thousand Oaks, CA: Sage Publications.

Temperley, D. (2007). *Music and probability*. Cambridge, MA: MIT Press.

Tillmann, B. (2012). Music and Language Perception: Expectations, Structural Integration, and Cognitive Sequencing. *Topics in Cognitive Science, 4*, 568–584.

Wolpert, D. M., & Flanagan, J. R. (2001). Motor prediction. *Current Biology*, *11*, R729-R732.

Youngblood, J. E. (1958). Style as information. *Journal of Music Theory*, 2, 24–35.

Appendix

Research suggests that musical expectation depends on incidental exposure to Western tonal music (Bharucha, 1987; Huron, 2006; Meyer, 1956), but it is unclear whether this prior musical knowledge contributes to expectation and memory for novel sequences that lack stylistically familiar structure. Framed more generally, how does previously encoded knowledge have an impact on statistical learning and memory representations for structurally unfamiliar stimuli?

To address this question, the simulations reported in Section 3 were compared with models pre-trained on a corpus of Western tonal music. The LTM+ was trained using a corpus of 905 Western folk songs and hymns (Bach chorale melodies, German folk songs and Canadian ballads) used in previous research (see Pearce & Wiggins, 2006, Table 2) to simulate, at a general level, the long-term schematic exposure of an average Western listener. We chose this corpus because models trained on it have been found to simulate accurately listeners' pitch expectations (Hansen & Pearce, 2014; Pearce, 2005; Pearce et al., 2010a; Omigie et al., 2012, 2013). Furthermore, this was the same corpus used to construct the prior for the model that generated the stimuli in the behavioral study.

We hypothesized that if listeners did not draw upon their schematic knowledge of Western tonal music, but instead acquired a new cognitive representation of the statistical structure of the stimuli through exposure, then performance on both tasks should be better simulated by BOTH+ models without pre-training (see Section 3) than those reported here in the Appendix. The simulation results are reported in Tables A1 (expectedness) and A2 (memory). Pre-training did not improve the accuracy of either the expectedness or memory simulations.

----------Insert Table A1 about here----------

The results suggest that listeners acquired a cognitive representation of statistical regularities in the stimulus set that was distinct from their existing representation of statistical structure in music. In terms of predictive coding (Clark, 2013; Friston, 2010), we suggest that the prediction error between the listener's top-down musical expectations and the ongoing stimulus undermines the utility of pre-existing predictive models, stimulating the construction of new predictive models through dynamic statistical learning within stimuli (in the STM) and from the entire stimulus set (in the LTM+).

----------Insert Table A2 about here----------

Figure captions

Fig. 1. The three-dimensional space from which stimuli were selected. Stimuli presented in both the listening and test sessions are shown, as well as the distractor stimuli heard only in test sessions and not listening sessions.

Fig. 2. Top: A sample tone sequence with moderately low Sequence Information Content (1.06 nats), low Sequence Predictive Information (0.31), high Sequence Coding Gain (0.83), and low Probe Tone Information Content (0.86). Bottom: A sample tone sequence with high Sequence Information Content (3.19), moderate Sequence Predictive Information (0.55), low Sequence Coding Gain (-1.11), and high Probe Tone Information Content (5.17). The probe tone in both sequences is marked with an arrow. Across all stimuli, Probe Tone Information Content values range from 0.04 to 8.44, Sequence Information Content values range from 0.24 to 3.43, Sequence Predictive Information values range from 0.04 to 0.78, and Sequence Coding Gain values range from -1.33 to 1.19.

Fig. 3. Probe Tone Information Content (in nats, where 1 nat = 1.44 bits) as a predictor of average expectedness ratings of probe tones.

Fig. 4. The differential effect of Sequence Predictive Information on Proportion Correct Response during recognition memory tests for New and Old sequences. Note that Proportion Correct Response is used here rather than the categorical variable Correct Response for clarity of illustration.

Table 1.

Relationship between Information-theoretic measures: Pointwise measures reflect a single event in a sequence, obtained by taking the double integral of the corresponding distribution measure, while sequence measures reflect the average of the corresponding pointwise measure across all events in a sequence.

Distribution measures	Pointwise measures	Sequence measures
Entropy Rate	Information Content	Sequence Information Content
Multi-Information Rate	Coding Gain	Sequence Coding Gain
Predictive Information Rate	Predictive Information	Sequence Predictive Information

Table 2.

Recognition memory test performance (proportion correct) for *Old* and *New* sequences across listening sessions.

Listening Session	Old/Familiar (Hits)	New/Unfamiliar (Correct Rejections)
Session 1	0.67	0.64
Session 2	0.63	0.65
Session 3	0.70	0.65

Table 3.

IDyOM simulations of expectedness ratings, showing correlation coefficients between expectedness and probe tone information content for each listening session individually (DF = 22) and across all three sessions (DF = 70).

Configuration	Order	Feature	Session 1	Session 2	Session 3	Overall correlation
LTM+	Zero	Pitch	-0.45	-0.16	-0.20	-0.29
LTM+	First	Pitch	-0.64	-0.58	-0.62	-0.61*
LTM+	Variable	Pitch	-0.66*	**-0.85***	-0.37*	-0.51*
BOTH+	Zero	Pitch	-0.78*	-0.79*	-0.80*	-0.79*
BOTH+	First	Pitch	-0.76*	-0.68*	-0.70*	-0.71*
BOTH+	Variable	Pitch	-0.71*	-0.69*	-0.45*	-0.57*
LTM+	Zero	Interval	-0.58	-0.45	-0.49	-0.50*
LTM+	First	Interval	-0.76*	-0.72*	-0.72*	-0.73*
LTM+	Variable	Interval	-0.78*	-0.76*	-0.60*	-0.67*
BOTH+	*Zero*	*Interval*	**-0.89***	**-0.85***	**-0.86***	***-0.86****
BOTH+	First	Interval	-0.84*	-0.82*	-0.83*	-0.83*
BOTH+	Variable	Interval	-0.81*	-0.77*	-0.65*	-0.71*

Note: * denotes $p < .001$ (Bonferroni corrected). Bold font indicates highest correlation in each column. Italic font indicates the model with the highest overall correlation.

.

Table 4.

IDyOM simulations of memory performance, showing correlation coefficients between D-prime score and sequence information content for each test session individually (DF = 6) and across all three sessions (DF = 22).

Configuration	Order	Feature	Session 1	Session 2	Session 3	Overall correlation
LTM+	Zero	Pitch	0.21	-0.30	-0.46	-0.20
LTM+	First	Pitch	-0.23	-0.65	-0.51	-0.48
LTM+	Variable	Pitch	-0.30	-0.69	-0.74	-0.34
BOTH+	Zero	Pitch	-0.15	-0.66	-0.75	-0.48
BOTH+	First	Pitch	-0.13	-0.71	-0.80	-0.56
BOTH+	Variable	Pitch	-0.36	-0.71	-0.83	-0.42
LTM+	Zero	Interval	-0.21	-0.61	-0.80	-0.53
LTM+	First	Interval	-0.27	-0.75	-0.88	-0.60
LTM+	Variable	Interval	-0.34	-0.76	-0.81	-0.41
BOTH+	Zero	Interval	-0.40	**-0.81**	**-0.92**	-0.70*
BOTH+	*First*	*Interval*	**-0.48**	**-0.81**	-0.91	***-0.72****
BOTH+	Variable	Interval	-0.36	-0.78	-0.84	-0.44

Note: * denotes $p < .001$ (Bonferroni corrected). Bold font indicates highest correlation in each column. Italic font indicates the model with the highest overall correlation.

Table A1.

Pre-trained IDyOM simulations of expectedness ratings, showing correlation coefficients between expectedness and probe tone information content for each listening session individually (DF = 22) and across all three sessions (DF = 70).

Configuration	Order	Feature	Session 1	Session 2	Session 3	Overall correlation
LTM+	Zero	Pitch	0.02	0.01	0.01	0.02
LTM+	First	Pitch	-0.43	-0.41	-0.41	-0.42*
LTM+	Variable	Pitch	-0.66*	**-0.85***	-0.47*	-0.52*
BOTH+	Zero	Pitch	-0.77*	-0.79*	-0.79*	-0.78*
BOTH+	First	Pitch	-0.58	-0.53	-0.53	-0.54*
BOTH+	Variable	Pitch	-0.75*	-0.76*	-0.59*	-0.62*
LTM+	Zero	Interval	-0.27	-0.24	-0.24	-0.25
LTM+	First	Interval	-0.37	-0.41	-0.4	-0.39*
LTM+	Variable	Interval	-0.57	-0.61	-0.54	-0.49*
BOTH+	*Zero*	*Interval*	**-0.88***	-0.83*	**-0.84***	***-0.85****
BOTH+	First	Interval	-0.64*	-0.67*	-0.69*	-0.66*
BOTH+	Variable	Interval	-0.67*	-0.7*	-0.61*	-0.58*

Note: * denotes $p < .001$ (Bonferroni corrected). Bold font indicates highest correlation in each column. Italic font indicates the model with the highest overall correlation.

Table A2.

Pre-trained IDyOM simulations of memory performance, showing correlation coefficients between D-prime score and sequence information content for each test session individually (DF = 6) and across all three sessions (DF = 22).

Configuration	Order	Feature	Session 1	Session 2	Session 3	Overall correlation
LTM+	Zero	Pitch	0.22	0.26	-0.13	0.09
LTM+	First	Pitch	-0.2	-0.52	-0.58	-0.43
LTM+	Variable	Pitch	-0.29	-0.73	-0.67	-0.3
BOTH+	Zero	Pitch	-0.12	-0.66	-0.79	-0.48
BOTH+	First	Pitch	-0.08	-0.63	-0.67	-0.47
BOTH+	Variable	Pitch	-0.35	**-0.81**	-0.72	-0.37
LTM+	Zero	Interval	-0.19	-0.38	-0.44	-0.35
LTM+	First	Interval	-0.09	-0.6	-0.68	-0.42
LTM+	Variable	Interval	-0.35	-0.47	-0.76	-0.31
BOTH+	*Zero*	*Interval*	-0.35	-0.76	**-0.91**	***-0.64****
BOTH+	First	Interval	-0.36	-0.62	-0.85	-0.61
BOTH+	Variable	Interval	**-0.43**	-0.63	-0.77	-0.39

Note: * denotes $p < .001$ (Bonferroni corrected). Bold font indicates highest correlation in each column. Italic font indicates the model with the highest overall correlation.